

Discovery of structurally diverse HIV-1 integrase inhibitors based on a chalcone pharmacophore

Jinxia Deng,^a Tino Sanchez,^a Laith Q. Al-Mawsawi,^a Raveendra Dayam,^a
Rosendo A. Yunes,^b Antonio Garofalo,^c Michael B. Bolger^d and Nouri Neamati^{a,*}

^aDepartment of Pharmacology and Pharmaceutical Sciences, University of Southern California, School of Pharmacy,
1985 Zonal Avenue, Los Angeles, CA 90089, USA

^bDepartamento de Química, Universidad Federal de Santa Catarina, Florianópolis, Santa Catarina, Brazil

^cDipartimento di Scienze Farmaceutiche, Università della Calabria, 87036 Arcavacata di Rende (Cs), Italy

^dSimulations Plus, Inc., 42505 10th Street West, Lancaster, CA 93543-2902, USA

Received 7 January 2007; revised 18 February 2007; accepted 15 April 2007

Available online 25 April 2007

Abstract—Recently, we reported small-molecule chalcones as a novel class of HIV-1 integrase (IN) inhibitors. The most potent compound showed an IC₅₀ value of 2 μM for both IN-mediated 3'-processing and strand transfer reactions. To further utilize the chalcones, we developed pharmacophore models to identify chemical signatures important for biological activity. The derived models were validated with a collection of published inhibitors, and then were applied to screen a subset of our small molecule database. We tested 71 compounds in an in vitro assay specific for IN enzymatic activity. Forty-four compounds showed inhibitory potency <100 μM, and four of them exhibited IC₅₀ values <10 μM. One compound, **62**, with an IC₅₀ value of 0.6 μM, displayed better potency than the original chalcone **2** against the strand transfer process. This study demonstrates the systematic use of pharmacophore technologies to discover novel structurally diverse inhibitors based on lead molecules that would exhibit poor characteristics in vivo. The identified compounds have the potential to exhibit favorable pharmacokinetic and pharmacodynamic profiles.
© 2007 Elsevier Ltd. All rights reserved.

1. Introduction

Integrase (IN) is one of the three essential enzymes encoded by the human immunodeficiency virus (HIV) *pol* gene. The other two enzymes, reverse transcriptase and protease, have been widely exploited as antiretroviral drug targets, and currently there are 21 drugs available for the treatment of HIV infected patients. However, due to the development of drug resistance and cytotoxic side effects, many patients experience unsatisfactory virologic, immunologic, or clinical outcomes from currently available therapies, limiting therapeutic options. These aspects require the identification of novel drugs targeting a different stage of the retroviral life cycle. HIV-1 IN has been validated as a potential target because of its essential role in viral replication. The exis-

tence of a rapid and sensitive in vitro inhibition assay is also very advantageous to identify novel lead compounds for future drug development. In addition, crystal and NMR structures are available to utilize for structure-based drug design.^{1–5} These features make IN a well-suited target for drug design with a low potential for mechanism-based toxicity in humans. Nevertheless, IN is a challenging system for structure-based drug discovery because of its shallow substrate binding site, the fact that it functions in a multimeric form during catalysis, and the lack of a full-length enzyme structure. Over the past years, a number of inhibitors have been reported to inhibit IN.^{6–19} Several compounds such as S-1360, L-870,810, MK-0518, and GS-9137 have thus far advanced to clinical trials.^{16,20–23}

Recently, we reported 11 structurally novel compounds showing antiviral activity in the NCI drug screening program by our enzyme based assays specific for IN.¹² Chalcones **1** and **2** were selected as potential leads (Fig. 1). Unfortunately, many chalcones are non-specific inhibitors and possess cytotoxicity in a large number of

Keywords: HIV integrase inhibitor; Chalcone; Drug design; Database searching; Pharmacophore model; ADME; Docking; eHiTS; GOLD.

* Corresponding author. Tel.: +1 323 442 2341; fax: +1 323 442 1390; e-mail: neamati@usc.edu

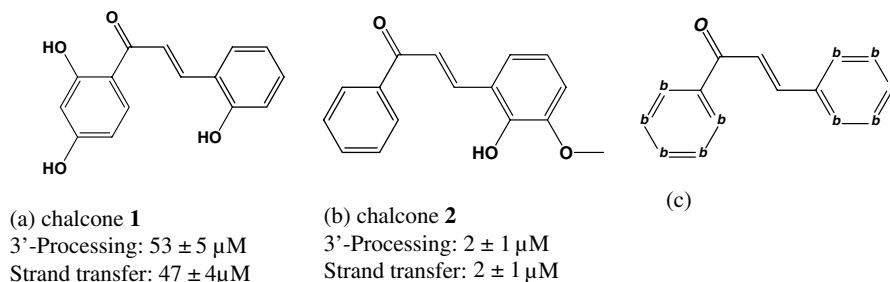


Figure 1. (a) and (b) Structures of the two recently reported chalcones. (c) The substructure query derived from the chalcones. 'b' represents atom C, N, or S, and the number of hydrogen atoms could be any.

tumor cell lines. In addition to anticancer properties,^{24–26} chalcones and analogues exhibited antibacterial, antifungal, and antiviral properties.²⁷

We therefore aimed to discover non-chalcone-based compounds using the previously identified chalcones as starting leads for the design of pharmacophore models. The pharmacophore-based technique is well suited to identify novel compounds with the same biological activity as the lead compound, but with diverse structural chemical scaffolds (for a recent review, please see Ref. 28). The identified compounds are expected to exert different physicochemical and pharmacokinetics properties from the starting molecules. In this work, we carried out pharmacophore model studies, including an initial substructure search, to explore the possible chemical features important for chalcone potency. The models were then used as search queries to screen a small molecule database. Selected compounds were tested for activity in IN-specific *in vitro* inhibition assays. Following this procedure, we identified several active compounds, with one of them displaying potent strand transfer inhibition in the high nanomolar range.

2. Methods

2.1. Substructure search

2-D substructures were built and converted to hypotheses by the Catalyst software package.²⁹ The initial hypothesis was derived from both chalcones **1** and **2** (Fig. 1), and then applied as a query to search a subset of our in-house small molecule database in order to identify chalcone-template containing molecules. The database contains approximately 360,000 small-molecule compounds. For each compound, a collection of up to 250 conformations were generated by Catalyst.

2.2. Functional feature pharmacophore model

The functional feature pharmacophore model was generated using a single conformation of either chalcones **1** or **2** by the Catalyst software package (Accelrys, Inc.). First, the desired functional features, that is, H-bond donor, H-bond acceptor, and hydrophobic feature, were mapped and added to the selected conformation, which is either the energy-minimized conformation or the best predicted conformation from docking calculations in this

study. The poling algorithm with CHARMM force field was then used to generate the energy-optimized conformation by the 'BEST conformation-generation routine' in Catalyst. Next, we assigned the constraints to each feature (e.g., coordinate and size of the feature). We chose the default value for each selected feature to avoid negative fitting values. Finally, all the selected features were merged into one pharmacophore model. The same procedure was then repeated for the other chalcone. The derived pharmacophore models were used to screen our in-house database to identify potential IN inhibitors without a chalcone-like chemical scaffold.

2.3. Docking simulations

The docking studies were carried out by the eHiTS docking program, a new, fast, and exhaustive flexible ligand docking system to complement our small molecule docking procedure.^{30,31} eHiTS evaluates all the possible protonation states for the receptor and ligands automatically for every receptor–ligand pair. The docking method systematically covers the part of the conformational and positional search space to avoid severe steric clashes.^{30,31} The program has been validated using 1626 protein–ligand complexes from the protein databank, which is the largest published validation set thus far. The validation complexes were divided into two subsets, one has 844 complexes used as the training set, and the remaining 742 complexes were used in a testing set. eHiTS yielded over 71% of the test set top-ranked structure complexes with a RMSD of less than 2.0 Å as compared with the experimentally derived X-ray conformation, while 75% of the complexes in the training set obtained similar results.^{30,31}

The eHiTS package is comprised of the main application, a set of utility programs, and a running script that automates and simplifies the conformational sampling procedure. We used the cocrystal structure of IN bound with the inhibitor 5CITEP complex (PDB 1QS4³) as the starting structure. Chain A, bound to both the inhibitor and a divalent metal cation (Mg^{2+} chelated D64 and D116), was selected as the docking target. The ligand (5CITEP) was subsequently removed to keep the binding pocket available for docking analysis. We set input files containing ligands and receptor while defining the docking surface. The receptor and ligands were then pre-processed. The fully automated pre-processing phase includes separation of receptor and ligand from

a PDB complex file, conversion of the input receptor and ligands into tagged molecular file format (native eHiTS file format), and generation of steric grid and feature graph description of the cavity's characteristics. Next, the eHiTS docking and scoring program was launched. eHiTS generated 3D coordinates of docked conformations of ligands in the receptor's cavity. The generated poses were scored and saved into a tagged molecular ascii (TMA) file format. The last step of the process was conversion of the docked poses from TMA to SDF files for conformational visualization and analysis.^{30,31}

To further train the eHiTS score to enhance the program performance by ranking active compounds over inactives, we applied the enhancement training module available with the eHiTS package. The module takes a set of active ligands, and an inactive set containing at least five-times more number of decoys than the number of the actives for analysis. We selected five published active molecules³² with IC_{50} values < 10 nM against the strand transfer process (see supporting information) and one more active compound from our own recent studies (unpublished data) to compose the active set. Forty-one compounds (unpublished data obtained from our laboratory) with IC_{50} values > 100 μ M were used in the decoy set.

Finally, we performed docking studies on the same set of newly tested compounds from this work by the GOLD (version 1.2 Genetic Optimization for Ligand Docking) software package³³ running on our Silicon Graphics Onyx workstation as described elsewhere.^{12,34}

3. HIV-1 IN inhibition assay

3.1. Materials, chemicals, and enzymes

All compounds were dissolved in DMSO, and the stock solutions were stored at -20°C . The $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ was purchased from either Amersham Biosciences or ICN. The expression system for the wild-type IN was a generous gift of Dr. Robert Craigie, Laboratory of Molecular Biology, NIDDK, NIH (Bethesda, MD).

3.2. Preparation of oligonucleotide substrates

The oligonucleotides 21top, 5'-GTGTGGAAATCTC TAGCAGT-3', and 21bot, 5'-ACTGCTAGAGATTT TCCACAC-3', were purchased from Norris Cancer Center Microsequencing Core Facility (University of Southern California) and purified by UV shadowing on polyacrylamide gel. To analyze the extent of 3'-processing and strand transfer using 5'-end-labeled substrates, 21top was 5'-end-labeled using T₄ polynucleotide kinase (Epicenter, Madison, WI) and $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ (Amersham Biosciences or ICN). The kinase was heat-inactivated, and 21bot was added in 1.5 M excess. The mixture was heated at 95°C , allowed to cool slowly to room temperature, and processed through a spin 25 minicolumn (USA Scientific, Ocala, FL) to separate annealed double-stranded oligonucleotide from unincorporated material.

3.3. IN assays

To determine the extent of 3'-processing and strand transfer, wild-type IN was preincubated at a final concentration of 200 nM with the inhibitor in reaction buffer [50 mM NaCl, 1 mM Hepes, pH 7.5, 50 μ M EDTA, 50 μ M dithiothreitol, 10% glycerol (w/v), 7.5 mM MnCl_2 , 0.1 mg/mL bovine serum albumin, 10 mM 2-mercaptoethanol, 10% dimethylsulfoxide, and 25 mM MOPS, pH 7.2] at 30°C for 30 min. Then, a 20 nM concentration of the 5'-end ^{32}P -labeled linear oligonucleotide substrate was added, and incubation was continued for an additional 1 h. Reactions were quenched by the addition of half the sample volume (8 μ L) of loading dye (98% deionized formamide, 10 mM EDTA, 0.025% xylene cyanol, and 0.025% bromophenol blue). An aliquot was electrophoresed on a denaturing 20% polyacrylamide gel (0.09 M Tris–borate, pH 8.3, 2 mM EDTA, 20% acrylamide, and 8 M urea).

Gels were dried, exposed in a PhosphorImager cassette, and analyzed using a Typhoon 8610 Variable Mode Imager (Amersham Biosciences, Piscataway, NJ) and quantitated using ImageQuant 5.2. The percent inhibition (% *I*) was calculated using the following equation

$$\%I = 100 \times [1 - (D - C)/(N - C)], \quad (1)$$

where *C*, *N*, and *D* are the fractions of 21-mer substrate converted to 19-mer (3'-processing product) or strand transfer products for DNA alone, DNA plus IN, and IN plus drug, respectively. The IC_{50} values were determined by plotting the logarithm of drug concentration versus percent inhibition to obtain the concentration that produced 50% inhibition.

4. Results and discussions

4.1. Substructure search

To determine if the anti-IN activity of chalcones **1** and **2** was predominately due to the structural scaffold of this class of compounds or the side chain substituents, we conducted an initial substructure search. We extracted the chalcone backbone as a query (shown in Fig. 1c) to search our small molecule database. The number of hydrogen atoms was set to 'any' in order to identify all compounds in the database containing the chalcone motif. Three hundred and seventy nine compounds met the requirement of the input query, and nine molecules were initially selected for inhibition assays. Table 1 gives the structures and the IC_{50} values for both 3'-processing and strand transfer IN catalytic activities. Compound **3** showed moderate strand transfer inhibition (IC_{50} value of $39 \pm 18 \mu\text{M}$) and two chalcones, **4** and **5**, exhibited weak potency against this process, while the remaining six compounds, **5a–5f**, were inactive. All these tested chalcones exhibited poor inhibition against 3'-processing catalysis. The substructure search indicates that the chalcone backbone itself is not enough for optimal inhibition of IN enzymatic activities, as the starting chalcone **2** displays potent inhibition ($IC_{50} = 2 \mu\text{M}$) of both IN processes. This result

Table 1. Inhibition of catalytic activities of HIV-1 integrase by a series of chalcones identified from substructure model search

Compound	Structure	IC ₅₀ (μM)	
		3'-Processing	Strand transfer
3		>100	39 ± 18
4		>100	89
5		>100	92 ± 13
5a		>100	>100
5b		>100	>100
5c		>100	>100
5d		>100	>100
5e		>100	>100
5f		>100	>100

demonstrates that the side chain substituents of chalcones **1** and **2** must play an important role for the compounds, potency, encouraging us to explore the chemical features of these chalcones through pharmacophore analysis.

4.2. Pharmacophore model development

Pharmacophore studies based on chalcones **1** and **2** (Figs. 1a and b) were conducted to investigate the ex-

panded chemical space of the compounds with various chemical feature combinations. Figure 2 shows the pharmacophore models generated based on the energy minimized conformation of chalcone **1**. The six possible positions that may contribute to receptor binding were labeled as F1–F6, and could be mapped with functional motifs as illustrated in 2-D feature mapping (Fig. 2a). These features include an H-bond donor, an H-bond acceptor, a hydrophobic feature, and/or an aromatic

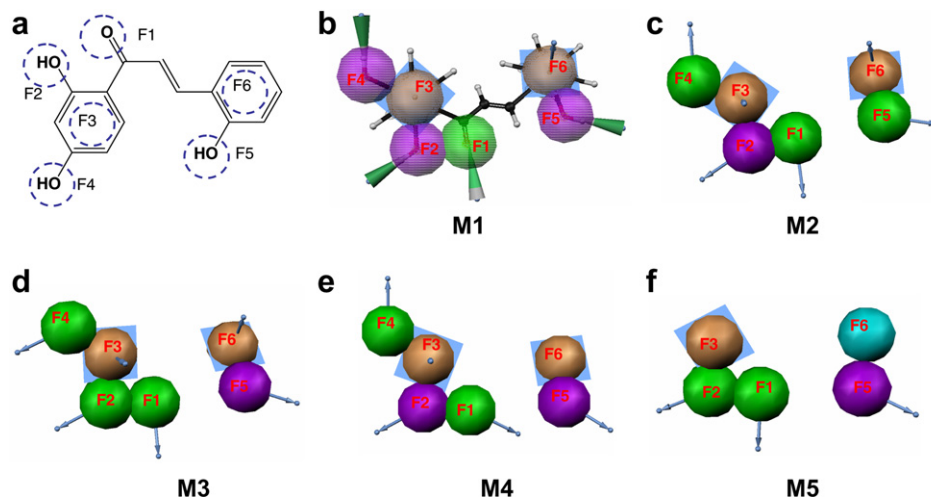


Figure 2. Five pharmacophore models generated from compound **1**. F1–F6 represents locations of the defined features. Green represents H-bond acceptor, magenta represents H-bond donor, brown represents aromatic ring feature, and light blue represents hydrophobic feature. (a) 2-D feature mapping against chalcone **1**. (b) An example of model (M1) mapping onto chalcone **1**. (c)–(f) Four independent pharmacophore models, M2, M3, M4, and M5.

ring feature. Figures 2b–f show the five possible pharmacophore models, labeled as M1, M2, M3, M4, and M5. Each model bears different feature combinations based on mapping of the template chalcone **1**. The variety originates from the chemical characteristics of the hydroxyl group, which could function as either an H-bond donor or an acceptor. Features that represent a hydroxyl group, that is, F2, F4, or F5, could therefore be assigned by either an H-bond donor or an H-bond acceptor, as shown in Figures 2b–e. The keto group feature F1, however, is always regarded as an H-bond acceptor group in all our models. Both the aliphatic carbon chain and the aromatic ring could be represented by a more general hydrophobic feature. Therefore, in the five models, F3 and F6 were all represented by the aromatic ring feature except in model M5, where the F6 position was mapped by a hydrophobic feature. Additionally, our docking studies suggested that the F4 hydroxyl group did not favorably interact with IN. Accordingly, we utilized this information during model development with the generation of model M5 (Fig. 2f) which contains only five features with the removal of the F4 hydroxyl group (Fig. 2d).

Using a similar approach, two four-feature pharmacophore models (Fig. 3) were derived from the most potent chalcone, **2**, which showed an IC_{50} value of $2 \pm 1 \mu M$ for both 3'-processing and strand transfer processes, respectively. The 2D feature mapping of the template is shown in Figure 3a, and the corresponding feature model based on the energy minimized conformation, M6, is displayed in Figure 3b. The four-feature bearing locations were labeled as P1–P4. P1 represents the keto group and is mapped as an H-bond acceptor, similar to the models derived from chalcone **1**. P3 represents the hydroxyl group and therefore could be mapped as either an H-bond acceptor or H-bond donor (presented is a model with P3 as an H-bond donor). P2 is mapped as a hydrophobic feature, and P4 is represented by an aromatic ring feature.

In addition, to investigate the conformational effect on model performance, we created a four-feature model, M7, in Figure 3c, using the most favorable orientation of chalcone **2** predicted from our docking calculations.¹² It is important to note that M7 and M6 represent different feature locations due to the difference of template conformations only.

5. Model validation

The derived models were validated by a subset of our small molecule database that contains a collection of both published known IN inhibitors and a set of compounds known to be inactive against IN. The active set contains 385 compounds with IC_{50} values less than $20 \mu M$ for either the 3'-processing or strand transfer process, or both. The inactive set is a collection of 235 compounds with IC_{50} values over $100 \mu M$ for both 3'-processing and strand transfer. Table 2 shows the performance of each presented model. Model selectivity is measured in terms of the selectivity index, SI, which is defined in Eq. 2,¹³

$$SI = \frac{\left(\frac{R_{active}}{R_{inactive}} \right)}{\left(\frac{D_{active}}{D_{inactive}} \right)}, \quad (2)$$

where

R_{active} is the total number of active inhibitors identified by the model;

$R_{inactive}$ is the total number of inactive compounds falsely identified by the model;

D_{active} is the total number of collected known active inhibitors in the database ($n = 385$);

$D_{inactive}$ is the total number of collected known inactive inhibitors in the database ($n = 235$).

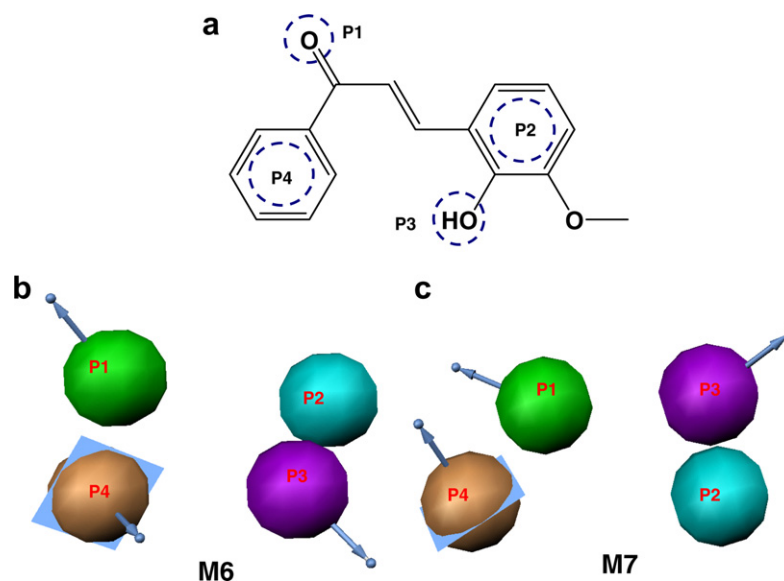


Figure 3. Feature models from chalcone **2**. P1–P4 represent locations of the defined features. H-bond acceptor is shown in green, H-bond donor in magenta, aromatic ring is in brown, and hydrophobic feature in cyan. (a) 2D feature mapping onto template chalcone **2**. (b) Pharmacophore model, M6, generated according to the energy-optimized conformation. (c) A pharmacophore model, M7, derived from the favorable orientation predicted by docking studies.

Table 2. Model validation by published IN inhibitors

Model	# of active ($n = 385$)	# of inactive ($n = 235$)	SI ^a
M1	69	12	3.50
M2	107	24	2.72
M3	129	26	3.02
M4	93	17	3.33
M5	188	72	2.61
M6	217	114	1.17
M7	241	105	1.40

^a Selectivity index, defined by Eq. 2.

The SI definition takes into account the size differences between the collected active and inactive inhibitor data sets. The six-feature models, M1–M4 (Figs 2b–e), were efficient at discriminating the active from the inactive compounds in terms of selectivity (SI: 2.72–3.50). The five feature model, M5 (Fig. 2f), showed less selectivity but recognized a higher rate of both active (48.9%) and inactive (30.6%) compounds over any of the six-feature models (M1–M4). The four-feature model, M6, with a SI of 1.17, could identify over 56% of all the active inhibitors, and nearly 49% of the inactive compounds. These results indicate that the number of model features plays a critical role in model selectivity while limiting the hit identification rate. However, model M7, derived from the favorable docked orientation, was able to recognize nearly 63% of the known active inhibitors and 45% of the inactive compounds with a SI value of 1.40, indicating that this model is more reliable than M6.

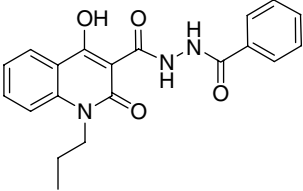
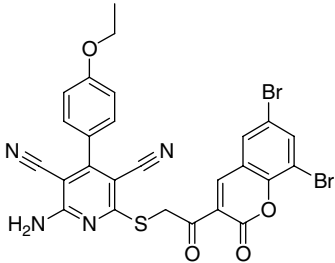
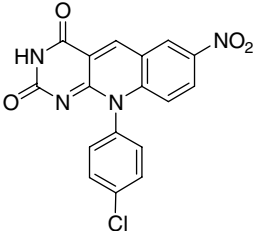
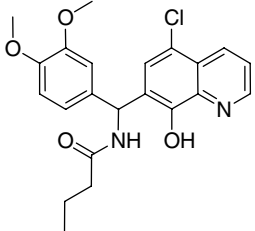
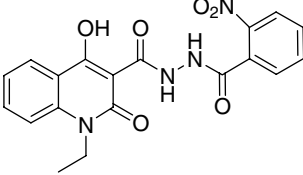
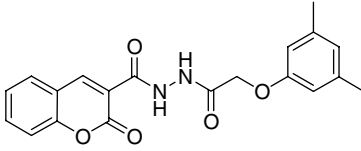
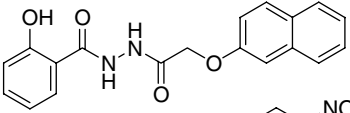
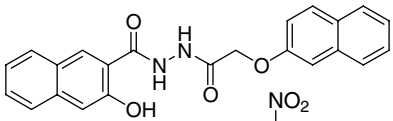
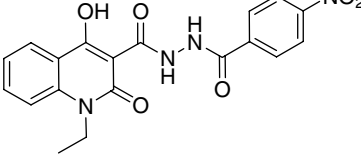
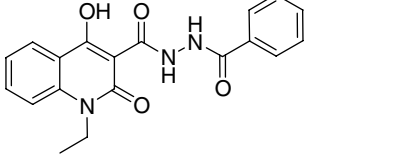
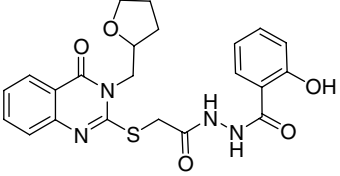
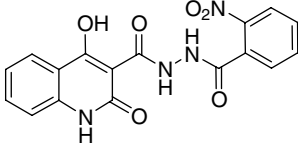
6. Database searching and IN inhibition assay

Based on our model validation results, we herein will only present the result of database mining using model M7 as the search query. A total of 407 compounds could

be mapped by model M7. Seventy-one of these compounds were tested against IN catalytic activity. Table 3 lists the structures of the tested compounds and their corresponding IC_{50} values. Forty-four compounds show inhibition potency at doses $<100 \mu M$. Compound **62** stands out as the most active compound in the set with an IC_{50} value of $1.9 \mu M$ for 3'-processing, and $0.6 \mu M$ for the strand transfer process, respectively. Other potent compounds, **7**, **18**, and **55**, all show a similar level of potency as chalcone **2**, with IC_{50} values below $10 \mu M$. **7** exhibits an IC_{50} value of $6 \mu M$ against 3' processing and $4 \mu M$ against the strand transfer process, respectively. This compound was further tested using Mg^{2+} as a cofactor. **7** exhibited a modest decrease in inhibition potency using Mg^{2+} with an IC_{50} value of $17 \mu M$ and $9.6 \mu M$ against 3'-processing and strand transfer, respectively. **18**, an analogue of earlier reported active compounds,³⁵ exhibited an IC_{50} value of $3 \mu M$ for the 3'-processing step and $1 \mu M$ for the strand transfer process, respectively.

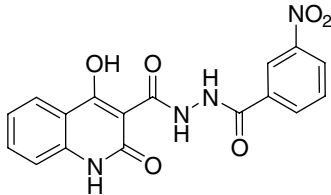
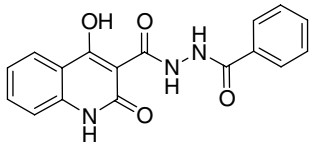
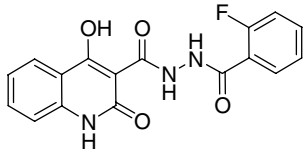
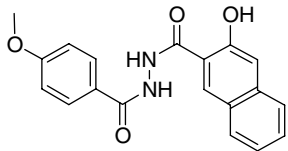
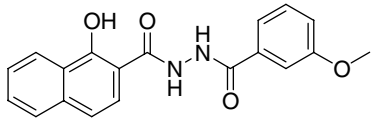
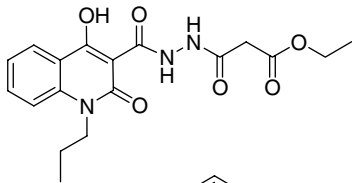
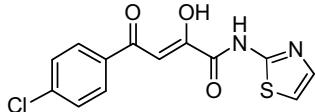
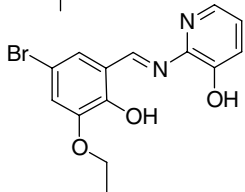
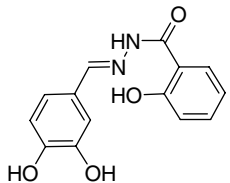
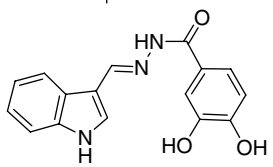
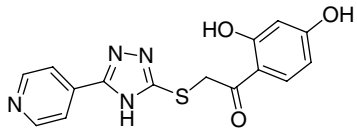
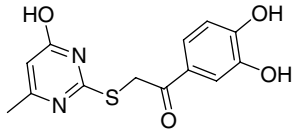
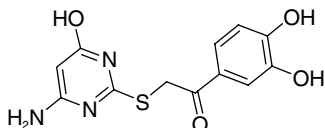
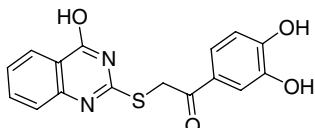
Shown in Figure 4 is an overlay of the mapping orientations of each novel compound (**62**, **7**, and **55**) over the pharmacophore model M7. Figure 4a depicts the orientation overlay of the most potent compound **62** onto M7, while Figure 4b illustrates that **7** could only partially fit the features of M7, but still retained desirable inhibitory activity. Figure 4c depicts the model fitting of **55**. Shown in Figure 4d is the superimposition of both **62** and chalcone **2** against M7, indicating that both molecules share a similar orientation by matching the phenol rings against the aromatic ring feature. The H-bond acceptor feature (green sphere in Fig. 4b) was mapped by the keto group of **2** and the nitrogen atom from **62**; the H-bond donor (magenta sphere in Fig. 4b) was mapped by the hydroxyl group from **2** and the NH group from **62**.

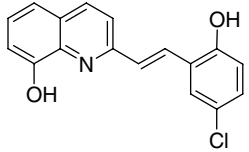
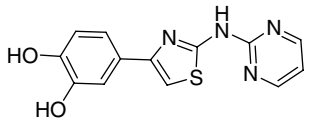
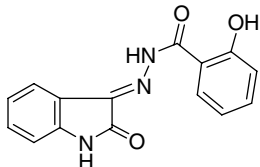
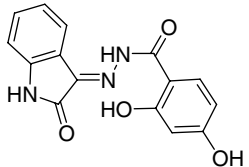
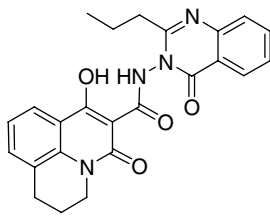
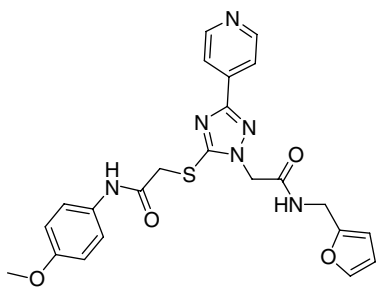
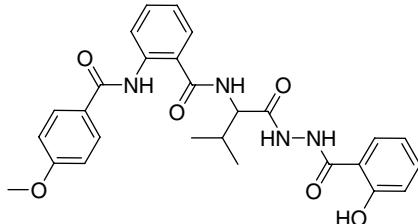
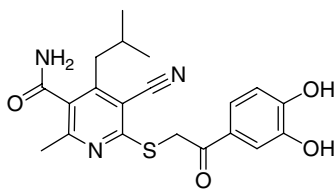
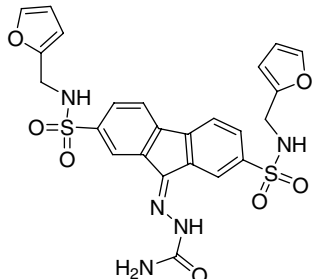
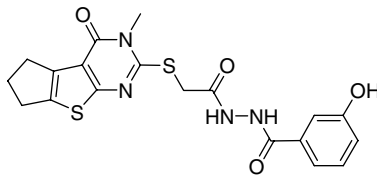
Table 3. Inhibition of HIV-1 integrase catalytic activities by hits identified from the feature model, M7^a

Compound	Structure	IC ₅₀ (μM)		Compound	Structure	IC ₅₀ (μM)	
		3'-Proc ^b	ST ^c			3'-Proc ^b	ST ^c
6		~100	66	7		6 ± 1 (17)	4 ± 3 (9.6)
8		78	49	9		>100	90
10		36	30	11		74	72
12		29	18	13		18	12
14		32	20	15		37	17
16		85	60	17		54	46

(continued on next page)

Table 3 (continued)

Compound	Structure	IC ₅₀ (μM)		Compound	Structure	IC ₅₀ (μM)	
		3'-Proc ^b	ST ^c			3'-Proc ^b	ST ^c
18		3.3	~1	19		~100	28
20		52	30	21		49	18
22		42	~33	23		>100	48
24		95 ± 9	70 ± 17	25		143 ± 51	81 ± 38
26		70 ± 26	26 ± 8	27		>1000	460
28		91 ± 16	25 ± 21	29		310	130 ± 52
30		278	89 ± 17	31		87 ± 23	77 ± 28

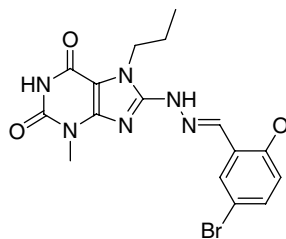
32		195	220	33		300	83 ± 20
34		>100	57 ± 20	35		>100	36 ± 17
36		64 ± 31	58 ± 9	37		>100	≥ 100
38		19 ± 1	16 ± 3	39		>100	85 ± 20
40		92 ± 16	20 ± 9	41		87	100

(continued on next page)

Table 3 (continued)

Compound	Structure	IC ₅₀ (μM)		Compound	Structure	IC ₅₀ (μM)	
		3'-Proc ^b	ST ^c			3'-Proc ^b	ST ^c
42		>100	>100	43		>100	>100
44		>100	>100	45		16	11
46		>100	>100	47		100	100
48		>100	>100	49		28	12

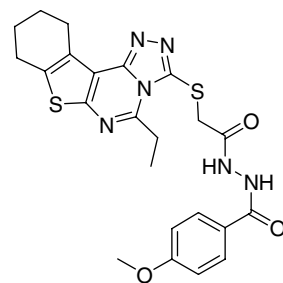
50



>100

80

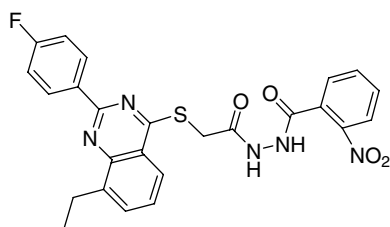
51



>100

100

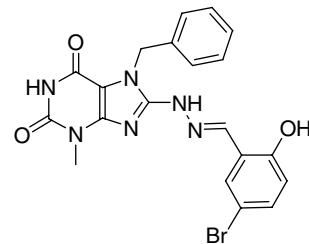
52



>100

100

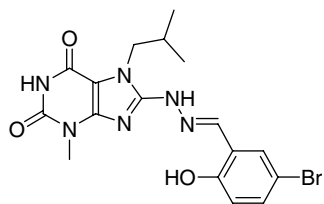
53



>100

>100

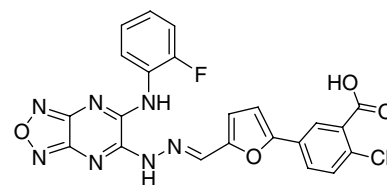
54



65

53

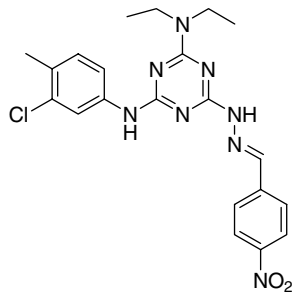
55



7

4

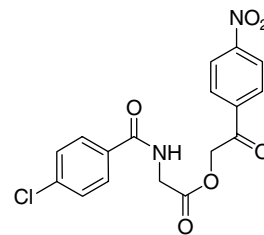
56



192

200

57

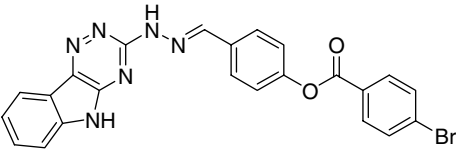
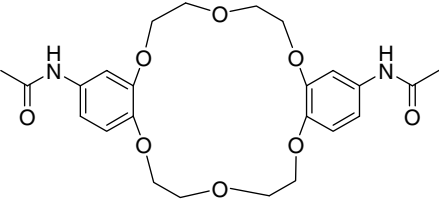
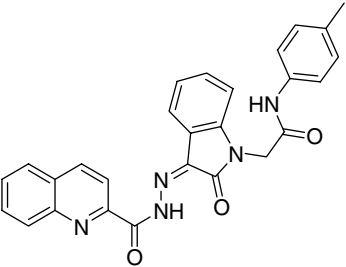
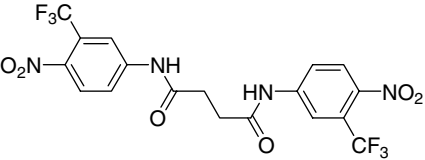
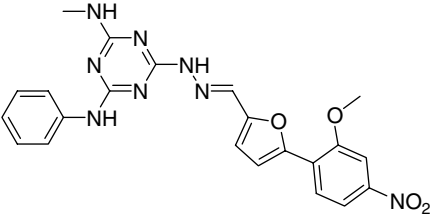
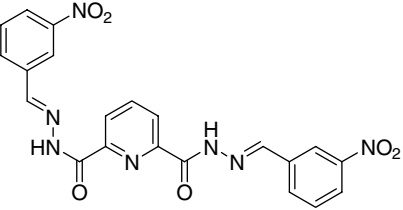
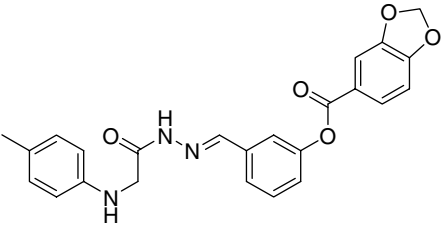
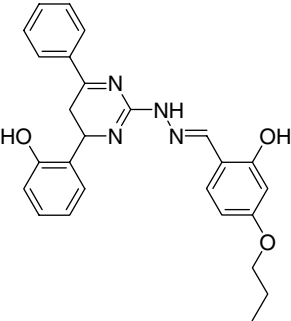
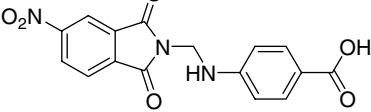


579 ± 156

542 ± 160

Table 3 (continued)

Compound	Structure	IC ₅₀ (μM)		Compound	Structure	IC ₅₀ (μM)	
		3-Proc ^b	ST ^c			3'-Proc ^b	ST ^c
58		67 74	32 100	59		138 282	189 200
60		>1000	>1000	61		43 29	33 29
62		1.9	0.6	63		161 448	251 439
64		240 212	246 224	65		80 197	70 208
66		18	13	67		41 54	44 43

68		857	800	69		679 1000	680 1000
70		84 ± 45	84 ± 40	71		196 ± 95	210 ± 109
72		81 ± 40	37 ± 10	73		150	130
74		1000 670	1000 731	75		64 ± 40	60 ± 33
76		349 180	234 180				

^a All compounds were initially screened at 1000 and 100 μ M, and only those that showed significant activity were subsequently tested 1–3 times at multiple lower doses to generate IC₅₀ values. IC₅₀ values were calculated from a plot of percentage of inhibition versus logarithm of concentration. Values in the parenthesis were obtained in the presence of Mg²⁺ as a cofactor.

^b 3'-Proc, 3'-processing.

^c ST, strand transfer reaction.

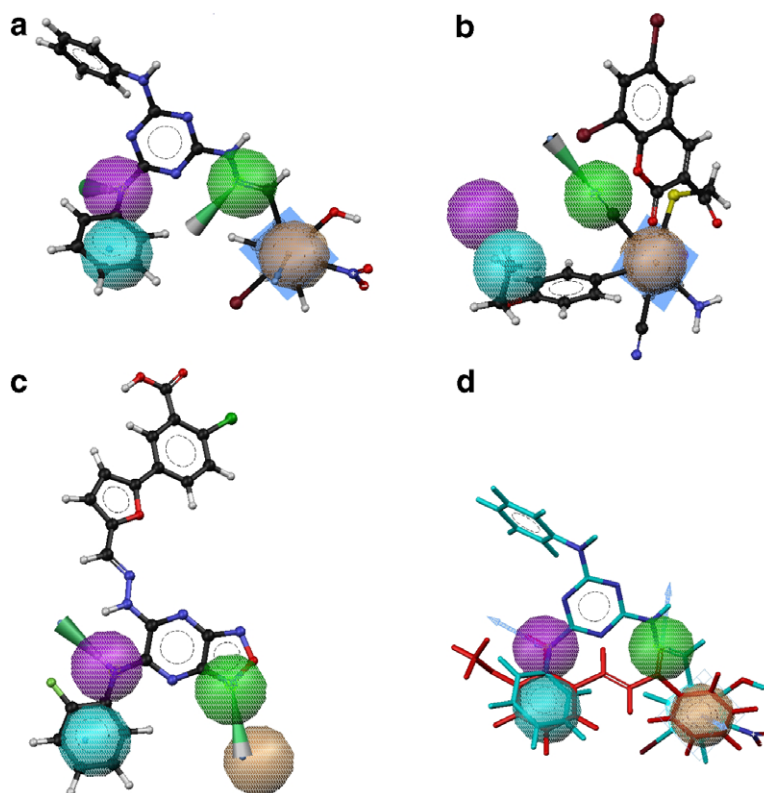


Figure 4. Selected compounds mapped against pharmacophore model M7. H-bond acceptor is shown in green, H-bond donor in magenta, hydrophobic feature in light blue, and aromatic ring in orange sphere. (a) **62**, (b) **7**, (c) **55**, (d) Superimposition of **62** (cyan) and chalcone **2** (red) against M7.

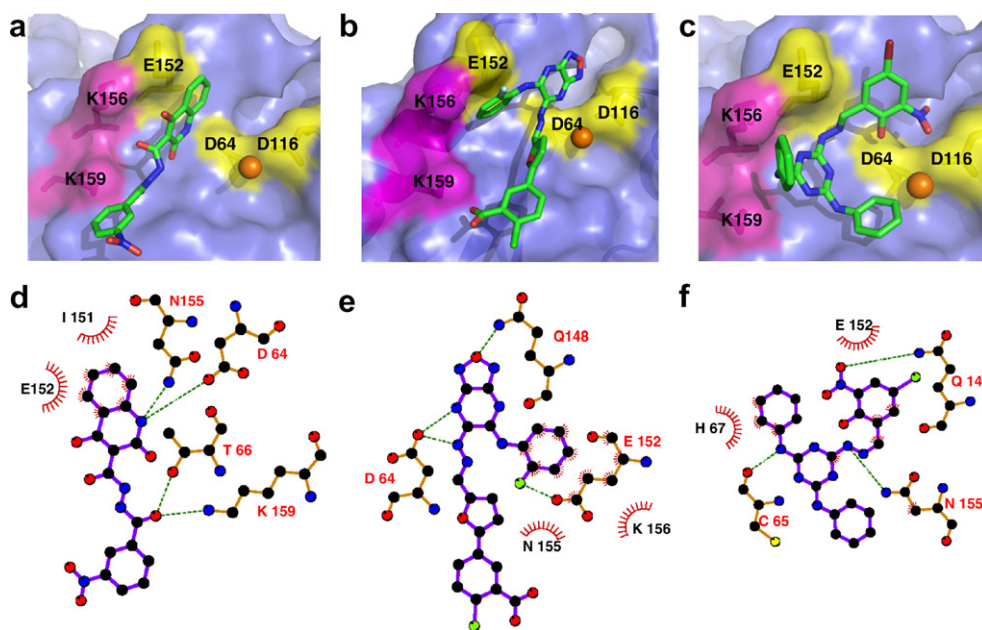


Figure 5. Predicted binding orientation of the representative novel inhibitors on the catalytic core domain of HIV-1 integrase. (a)–(c) show the ligand, **18**, **55**, and **62**, respectively, in the active site. Ligand is shown as stick model and binding surface is shown in blue. The brown sphere represents metal ion (Mg^{2+}). The images were generated by PyMOL. (d)–(f) demonstrate the detailed interactions between ligand, **18**, **55**, and **62**, respectively, with integrase in 2D plot representation, generated by LigPlot.

Table 4. Calculated properties of tested compounds^a

Compound	M.W.	Fitting value	eHiTS docking score	GOLD docking score	log <i>P</i>	#HBD	#HBA	#RB
6	365.4	2.97	−28.95	37.3	2.84	3	6	5
7	640.3	2.97	−28.86	65.8	1.96	1	8	5
8	368.7	1.99	−44.77	40.1	3.08	1	5	0
9	414.9	2.95	−35.27	48.8	2.78	2	6	8
10	396.4	0.42	−29.91	45.4	2.49	3	8	4
11	366.4	2.72	−36.71	60.1	2.57	2	7	6
12	336.3	3.24	−36.31	50.5	3.16	3	6	6
13	386.4	3.40	−48.05	66.3	3.83	3	6	6
14	396.4	0.20	−54.54	40.5	2.49	3	8	4
15	396.4	0.20	−31.59	41.1	2.49	3	8	4
16	454.5	3.19	−43.48	52.4	1.93	3	8	8
17	368.3	2.97	−56.11	47.6	2.02	4	8	3
18	368.3	2.97	−61.63	47.2	2.02	4	8	3
19	323.3	2.97	−40.92	43.5	2.14	4	6	3
20	341.3	2.97	−15.89	43.3	2.26	4	6	3
21	336.3	2.40	−31.96	49.3	3.43	3	6	4
22	336.3	2.65	−43.13	46.7	3.43	3	6	4
23	375.4	2.93	−34.93	47.9	1.18	3	8	9
24	308.7	1.68	−40.71	56.0	1.66	2	5	2
25	337.2	3.09	−40.43	60.6	1.98	2	5	2
26	272.3	2.49	−34.12	49.4	1.95	4	6	2
27	295.3	2.67	−48.05	54.8	1.92	4	5	2
28	328.3	3.75	−40.36	52.9	1.39	3	6	3
29	292.3	3.70	−48.98	58.4	0.97	3	6	3
30	293.3	3.71	−48.37	56.5	0.35	4	7	3
31	328.3	3.67	−48.12	58.9	1.59	3	6	3
32	297.7	3.12	−36.11	54.5	3.35	2	3	0
33	286.3	1.91	−36.75	51.2	1.70	3	6	2
34	281.3	3.58	−47.59	45.4	2.15	3	6	2
35	297.3	3.82	−43.14	45.5	1.39	4	7	2
36	430.5	3.53	−31.9	41.3	3.15	2	6	4
37	478.5	3.81	−33.58	62.9	0.76	2	9	11
38	504.5	3.74	−60.28	51.7	2.76	5	10	10
39	399.5	3.79	−39.79	48.8	1.36	3	7	5
40	555.6	3.89	−62.53	69.9	0.75	4	12	10
41	430.5	2.32	−31.96	54.3	1.49	3	7	6
42	395.4	3.03	−29.44	54.7	2.31	3	7	8
43	620.7	3.46	−32.41	61.8	1.40	2	11	13
44	446.5	3.52	−35.66	59.0	2.21	2	8	9
45	444.2	3.49	−37.59	13.9	3.08	3	9	4
46	475.3	3.50	−45.29	60.5	1.32	3	6	9
47	493.5	3.03	−53.26	40.4	1.77	4	9	5
48	460.5	3.80	−32.94	41.8	3.57	2	6	8
49	588.7	3.62	−39.42	69.0	−1.27	3	12	18
50	421.2	2.04	−37.49	60.9	2.80	3	6	4
51	496.6	2.71	−39.07	61.4	2.51	2	8	8
52	505.5	3.08	−46.28	51.2	3.98	2	8	7
53	469.3	3.03	−40.69	67.3	3.50	3	6	4
54	435.3	3.08	−40.26	62.4	3.04	3	6	4
55	493.8	2.83	−45.33	55.1	2.82	3	11	4
56	454.9	3.46	−54.22	8.3	5.16	2	9	7
57	376.7	2.62	−45.21	48.5	2.54	1	7	6
58	404.4	1.15	−31.53	38.8	2.38	4	8	4
59	341.3	2.52	−30.24	44.9	0.48	2	7	3
60	339.3	1.75	−37.5	53.7	5.02	1	4	6
61	472.5	3.74	−35.71	−32.5	4.23	4	11	7
62	521.3	3.79	−40.11	−27.6	5.05	4	10	6
63	654.4	3.15	−26.23	49.2	4.21	2	6	8
64	349.3	0.93	−31.29	57.7	2.34	2	8	5
65	426.4	3.39	−39.97	39.7	1.65	3	8	6
66	494.3	3.65	−40.43	17.9	3.97	3	9	8
67	342.3	2.96	−35.18	46.2	0.53	4	8	4
68	487.3	1.72	−26.24	58.9	3.90	2	7	4
69	474.5	3.01	−34.18	37.1	−0.02	2	10	4
70	463.5	2.98	−33.25	57.8	3.29	2	8	6

(continued on next page)

Table 4 (continued)

Compound	M.W.	Fitting value	eHiTS docking score	GOLD docking score	log <i>P</i>	#HBD	#HBA	#RB
71	494.3	3.25	−38.32	26.5	3.92	2	8	9
72	460.4	3.60	−45.12	20.5	4.08	3	11	6
73	461.4	0.82	−47.11	60.9	3.15	2	11	4
74	431.4	3.78	−28.22	51.4	2.38	2	8	7
75	442.5	3.80	−43.12	63.3	4.05	3	7	6
76	341.3	2.52	−30.24	45.0	0.48	2	7	3

^a Fitting values were obtained from mapping of compounds against the four-feature pharmacophore model, M7; docking scores were calculated using eHiTS and GOLD software packages. Lower eHiTS scores or higher GOLD scores indicate more favorable binding. log *P*, partition coefficient; #HBD, total number of H-bond donor; #HBA, total number of H-bond acceptor; #RB, total number of rotatable bonds, were calculated by ADMET Predictor software (Simulations Plus, Inc.).

Table 5. Ranking parameters analysis

	IC ₅₀ ≤ 20 μM		20 μM < IC ₅₀ ≤ 100 μM		IC ₅₀ > 100 μM		Enhancement factor ^c	
	3'-Proc	ST	3'-Proc	ST	3'-Proc	ST	3'-Proc	ST
Number of compounds ^a	8	14	32	34	31	23	1.00	1.00
GOLD score (≥ 50.5) ^b	4	7	13	15	19	14	0.82	0.82
eHiTS score ^c (≤ −38.3)	6	9	15	18	15	9	1.55	1.64
Fitting value (≥ 3.08) ^d	6	9	12	14	18	13	1.29	1.14

^a Total number of compounds in each subset.

^b Total number of compounds in the subset that could be ranked by the GOLD score with a cutoff value of 50.5. Thirty-six out of 71 compounds have GOLD score of ≥ 50.5.

^c Total number of compounds in the subset that could be ranked by the eHiTS score with a cutoff value of ≤ −38.3. The eHiTS score of −38.3 was selected since it yielded 36 out of 71 compounds as compared to GOLD for ranking purposes.

^d Total number of compounds in the subset could be ranked by the fitting value cut off using pharmacophore model M7. The fitting value of 3.08 was selected since it again yielded 36 out of 71 compounds as above for ranking purposes.

^e Enhancement factor is defined by Eq. 3.

7. Possible binding mode predictions

To investigate the possible binding mechanism, docking studies were performed on the tested compounds. The docking scores from both eHiTS and GOLD and the model fitting values together with other calculated properties of all compounds are summarized in Table 4. The eHiTS score in the table is rescaled due to the enrichment procedure. Selected physicochemical properties were calculated by the ADMET Predictor software package (Simulations Plus, Inc.). Most of the compounds satisfy Lipinski's rule of five (molecular weight < 500, partition coefficient log *P* < 5, number of H-bond donor < 5, and number of H-bond acceptor < 10). Compound 62 slightly deviated from these rules but nonetheless demonstrates the best anti-IN activity in the set.

Figure 5 depicts the feasible drug binding modes of 18, 55, and 62. The predicted orientations indicate that these compounds make favorable contacts in the substrate binding site. The most potent compound, 62, was predicted to favorably interact with IN by eHiTS as shown in both Figures 5c, f and Table 4. However the GOLD score failed to rank this compound as a well-binding molecule (Table 4). Compound 18 yielded the 2nd-top docking score of −61.63 from eHiTS (Table 4) due to the strong hydrophobic and H-bond interactions with neighboring residues of IN (Figs. 5a and d). One of the key residues in the active site, D64, and residue N155 are both observed to contribute an H-bond contact with 18.

K159, a critical residue for mononucleotide and DNA crosslinking,^{36,37} and T66 are also predicted to form an H-bond contact with 18. Compound 55 favorably interacts with IN as indicated by both a favorable score of −45.33 from eHiTS and 55.1 from GOLD. The 2D plot (Fig. 5e) based on the docking calculations indicates that D64, E152, and Q148 form H-bond contacts with 55 at different locations, and two other residues, K156 and N155, both have hydrophobic interactions with 55.

In addition, to compare the contribution of fitting values and docking scores as an efficient ranking factor for further hit identification, we divided our compounds into three subgroups. They were defined as active, moderately active, and inactive based on their IC₅₀ values. Table 5 shows the result of this analysis. In the active set, 14 compounds out of the total 71 have an IC₅₀ value < 20 μM for only the strand transfer process, while 8 compounds are active against 3'-processing below 20 μM. A parameter termed the enhancement factor (EF) is defined by Eq. 3,

$$EF = \frac{\left(\frac{S_{\text{inactive}}}{S_{\text{active}}}\right)}{\left(\frac{P_{\text{inactive}}}{P_{\text{active}}}\right)}, \quad (3)$$

where

*S*_{inactive} is the total number of inactive inhibitors (IC₅₀ > 100 μM) reported by the model, that is, 31, for 3'-processing subcategory, and 23 for strand transfer subcategory;

S_{active} is the total number of active compounds ($IC_{50} < 20 \mu\text{M}$) reported by the model, that is, 8 for 3'-processing subcategory, and 14 for strand transfer subcategory;

P_{inactive} is the total number of inactive inhibitors falsely ranked by the user-defined cutoff. For example, if we select a score value of 50.5 as the GOLD score cutoff, then 19 compounds have a score >50.5 , but indeed are inactive compounds against 3'-processing, that is, $P_{\text{inactive}} = 19$;

P_{active} is the total number of active inhibitors correctly ranked by the user-defined cutoff. For the same example above, with a GOLD score value of 50.5 as the cutoff, four active compounds against 3'-processing could be accurately ranked with a score >50.5 , that is, $P_{\text{active}} = 4$; therefore, the EF in the case of the GOLD prediction for 3'-processing subset is 0.82.

In a similar fashion, using the same number of top-ranked compounds sorted by the given cutoffs (GOLD score of 50.5, eHiTS score of -38.3 , and fitting value of 3.08, see Table 5), the enriched eHiTS score is the most efficient index to distinguish active over inactive compounds, compared with both the fitting value and GOLD score.

8. Conclusions

We carried out an extensive pharmacophore model analysis using two lead chalcones to investigate the possible chemical signatures and to discover novel IN inhibitors with diverse structural scaffolds. The pharmacophore models were generated based on either an energy minimized conformation or a docking-predicted orientation. The pharmacophore model based on the docking calculated conformation of chalcone **2** performed better than that upon energy-optimized conformation of the same compound. This pharmacophore model was successfully applied to identify a promising compound (**62**) exhibiting greater potency than the original chalcone. Three more compounds (**7**, **18**, and **55**) were identified exhibiting similar potency as the original chalcone. These results highlight the effectiveness of pharmacophore model development and the importance of its validation in the discovery of novel compounds with desired biological activity with diverse chemical scaffolds.

Acknowledgments

This study was supported by a grant from the Campbell Foundation. We thank L.D. Chiaradia and P.C. Leal for their technical support.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.bmc.2007.04.041](https://doi.org/10.1016/j.bmc.2007.04.041).

References and notes

- Dyda, F.; Hickman, A. B.; Jenkins, T. M.; Engelman, A.; Craigie, R.; Davies, D. R. *Science* **1994**, *266*, 1981.
- Goldgur, Y.; Dyda, F.; Hickman, A. B.; Jenkins, T. M.; Craigie, R.; Davies, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9150.
- Goldgur, Y.; Craigie, R.; Cohen, G. H.; Fujiwara, T.; Yoshinaga, T.; Fujishita, T.; Sugimoto, H.; Endo, T.; Murai, H.; Davies, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13040.
- Molteni, V.; Greenwald, J.; Rhodes, D.; Hwang, Y.; Kwiatkowski, W.; Bushman, F. D.; Siegel, J. S.; Choe, S. *Acta Crystallogr. D Biol. Crystallogr.* **2001**, *57*, 536.
- Cherepanov, P.; Sun, Z. Y.; Rahman, S.; Maertens, G.; Wagner, G.; Engelman, A. *Nat. Struct. Mol. Biol.* **2005**, *12*, 526.
- Kawasuji, T.; Yoshinaga, T.; Sato, A.; Yodo, M.; Fujiwara, T.; Kiyama, R. *Bioorg. Med. Chem.* **2006**, *14*, 8430.
- Dayam, R.; Deng, J.; Neamati, N. *Med. Res. Rev.* **2006**, *26*, 271.
- Dayam, R.; Neamati, N. *Curr. Pharm. Des.* **2003**, *9*, 1789.
- Neamati, N. *Expert Opin. Investig. Drugs* **2001**, *10*, 1767.
- Carlson, H. A.; Masukawa, K. M.; Rubins, K.; Bushman, F. D.; Jorgensen, W. L.; Lins, R. D.; Briggs, J. M.; McCammon, J. A. *J. Med. Chem.* **2000**, *43*, 2100.
- Nicklaus, M. C.; Neamati, N.; Hong, H.; Mazumder, A.; Sunder, S.; Chen, J.; Milne, G. W.; Pommier, Y. *J. Med. Chem.* **1997**, *40*, 920.
- Deng, J.; Kelley, J. A.; Barchi, J. J.; Sanchez, T.; Dayam, R.; Pommier, Y.; Neamati, N. *Bioorg. Med. Chem.* **2006**, *14*, 3785.
- Deng, J.; Sanchez, T.; Neamati, N.; Briggs, J. M. *J. Med. Chem.* **2006**, *49*, 1684.
- Deng, J.; Lee, K. W.; Sanchez, T.; Cui, M.; Neamati, N.; Briggs, J. M. *J. Med. Chem.* **2005**, *48*, 1496.
- Jin, H.; Cai, R. Z.; Schacherer, L.; Jabri, S.; Tsiang, M.; Fardis, M.; Chen, X.; Chen, J. M.; Kim, C. U. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 3989.
- Sato, M.; Motomura, T.; Aramaki, H.; Matsuda, T.; Yamashita, M.; Ito, Y.; Kawakami, H.; Matsuzaki, Y.; Watanabe, W.; Yamataka, K.; Ikeda, S.; Kodama, E.; Matsuoka, M.; Shinkai, H. *J. Med. Chem.* **2006**, *49*, 1506.
- Di Santo, R.; Costi, R.; Roux, A.; Artico, M.; Lavecchia, A.; Marinelli, L.; Novellino, E.; Palmisano, L.; Andreotti, M.; Amici, R.; Galluzzo, C. M.; Nencioni, L.; Palamara, A. T.; Pommier, Y.; Marchand, C. *J. Med. Chem.* **2006**, *49*, 1939.
- Walker, M. A.; Johnson, T.; Ma, Z.; Banville, J.; Remillard, R.; Kim, O.; Zhang, Y.; Staab, A.; Wong, H.; Torri, A.; Samanta, H.; Lin, Z.; Deminie, C.; Terry, B.; Krystal, M.; Meanwell, N. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 2920.
- Charvat, T. T.; Lee, D. J.; Robinson, W. E.; Chamberlin, A. R. *Bioorg. Med. Chem.* **2006**, *14*, 4552.
- Young, S. *The XIV International AIDS Conference*, Barcelona, Spain, 2002.
- Yoshinaga, Y.; Sato, A.; Fujishita, T.; Fujiwara, T. In *9th Conference on Retroviruses and Opportunistic Infections*, Seattle, WA, USA, 2002.
- Markowitz, M.; Nguyen, B.-Y.; Gotuzzo, F.; Mendo, F.; Ratanasuwana, W.; Kovacs, C.; Zhao, J.; Gilde, L.; Isaacs, R.; Teppler, H. *XVI International AIDS Conference*, Toronto Canada, 2006.
- DeJesus, E.; Berger, D.; Markowitz, M.; Cohen, C.; Hawkins, T.; Ruane, P.; Elion, R.; Farthing, C.; Cheng, A.; Kearney, B. *13th Conference on Retroviruses and Opportunistic Infections*, Denver, Colorado, 2006.

24. Kumar, S. K.; Hager, E.; Pettit, C.; Gurulingappa, H.; Davidson, N. E.; Khan, S. R. *J. Med. Chem.* **2003**, *46*, 2813.
25. Modzelewska, A.; Pettit, C.; Achanta, G.; Davidson, N. E.; Huang, P.; Khan, S. R. *Bioorg. Med. Chem.* **2006**, *14*, 3491.
26. Achanta, G.; Modzelewska, A.; Feng, L.; Khan, S. R.; Huang, P. *Mol. Pharmacol.* **2006**, *70*, 426.
27. Dimmock, J. R.; Elias, D. W.; Beazely, M. A.; Kandepu, N. M. *Curr. Med. Chem.* **1999**, *6*, 1125.
28. Deng, J.; Dayam, R.; Al-Mawsawi, L. Q.; Neamati, N. *Curr. Pharm. Des.* **2007**, *13*, 129.
29. Catalyst; 4.10 ed.; Accelrys Inc.: San Diego, CA.
30. Zsoldos, Z.; Reid, D.; Simon, A.; Sadjad, B. S.; Johnson, A. P. *Curr. Protein Pept. Sci.* **2006**, *7*, 421.
31. Zsoldos, Z.; Reid, D.; Simon, A.; Sadjad, S. B.; Johnson, A. P. *J. Mol. Graphics Modell.* **2006**.
32. Egbertson, M. S.; Moritz, H. M.; Melamed, J. Y.; Han, W.; Perlow, D. S.; Kuo, M. S.; Embrey, M.; Vacca, J. P.; Zrada, M. M.; Cortes, A. R.; Wallace, A.; Leonard, Y.; Hazuda, D. J.; Miller, M. D.; Felock, P. J.; Stillmock, K. A.; Witmer, M. V.; Schleif, W.; Gabryelski, L. J.; Moyer, G.; Ellis, J. D.; Jin, L.; Xu, W.; Braun, M. P.; Kassahun, K.; Tsou, N. N.; Young, S. D. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 1392.
33. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727.
34. Long, Y. Q.; Jiang, X. H.; Dayam, R.; Sanchez, T.; Shoemaker, R.; Sei, S.; Neamati, N. *J. Med. Chem.* **2004**, *47*, 2561.
35. Dayam, R.; Sanchez, T.; Neamati, N. *ChemMedChem* **2006**, *1*, 238.
36. Jenkins, T. M.; Esposito, D.; Engelman, A.; Craigie, R. *EMBO J.* **1997**, *16*, 6849.
37. Drake, R. R.; Neamati, N.; Hong, H.; Pilon, A. A.; Sunthakar, P.; Hume, S. D.; Milne, G. W.; Pommier, Y. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 4170.